

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

The role of counterfactual reasoning in responsibility judgments

Permalink

<https://escholarship.org/uc/item/7p6400p0>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43)

Authors

Wu, Sarah

Gerstenberg, Tobias

Publication Date

2021

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

The role of counterfactual reasoning in responsibility judgments

Sarah Wu

Stanford University, Stanford, California, United States

Tobias Gerstenberg

Stanford University, Stanford, California, United States

Abstract

To hold someone responsible, we need to assess what causal role their action played in bringing about the outcome. Causality, in turn, can be understood in terms of difference-making – C caused E if E would have been different had C been different. In previous work, the counterfactual of “C being different” is often construed as “C being absent”. Here, we explore how in social situations, this counterfactual can be alternatively construed by imagining how a different person would have acted in the same situation. We propose a computational model that formalizes this idea of counterfactual replacement. The model compares what actually happened with what would have happened had the person of interest been replaced by someone else, and then predicts responsibility to the extent that the replacement person would likely have changed the outcome. We test our model against people’s responsibility judgments in a variety of scenarios across several experiments.